Semidefinite programming and polynomials

Javier Peña, Carnegie Mellon Univ

joint work with
Juan Vera, Carnegie Mellon Univ
Luis Zuluaga, Univ of New Brunswick
Raphael Hauser, Oxford Univ

March 31, 2006
Tutorial on SDP, Uniandes/Externado

Plan

Semidefinite programming (SDP) and sums of squares (SOS)

Positive polynomials and SOS

Copositive matrices

Structured SDP

Beyond SDP

Semidefinite programming (SDP)

SDP: generalization of linear programming (LP).

$$\begin{array}{ll} \min & C \bullet X \\ \text{s.t.} & A_j \bullet X = b_j,\ j = 1, \ldots, m \\ & X \succeq 0. \end{array}$$

Here $C, A_j, X \in \mathbb{S}^n :=$ symmetric $n \times n$ matrices.

$M \succeq 0 \Leftrightarrow M \in \mathbb{S}^n_+ := \{M \in \mathbb{S}^n : x^{\mathrm{T}} M x \geq 0 \ \ \forall x \in \mathbb{R}^n\}$

$X \bullet S := \mathrm{trace}(XS) = \sum_{ij} X_{ij} S_{ij}$.

Nesterov & Nemirovskii, Alizadeh 1980s & 1990s:
Interior-point methods (ipm) for SDP.

Applications: control, combinatorial optimization,...

Natural connection SDP/sums of squares (SOS):

$M \succeq 0 \Leftrightarrow M = LL^{\mathrm{T}}$ for some $L \in \mathbb{R}^{n \times n}$ (Cholesky factorization)
Can rephrase $M = LL^{\mathrm{T}}$ as

$$x^{\mathrm{T}} M x = \sum (l_i^{\mathrm{T}} x)^2,$$

where $L = \begin{bmatrix} l_1 & \cdots & l_n \end{bmatrix}$.

In other words, $M \succeq 0 \Leftrightarrow x^{\mathrm{T}} M x$ is a sum of squares of linear forms.

Positive polynomials and sums of squares (SOS)

Suppose p, q_i are polynomials in n variables.
Observe: If $p = \sum_i q_i^2$ then $p(x) \geq 0$ for all $x \in \mathbb{R}^n$.

Theorem (Hilbert 1800s)

Converse is true if and only if

$\deg(p) = 2$,

or $n = 1$,

or $n = 2$ *&* $deg(p) = 4$.

Say that p is a sum of squares (SOS) if $p = \sum_i q_i^2$ for some polynomials q_i.

SOS can be naturally phrased in SDP terms.

Given a polynomial q can write $q(x) = c(q)^{\mathrm{T}} v(x)$
where $c(q)$: vector of coefficients of $q(x)$
$v(x)$: vector of monomials in x of degree $\leq \deg(q)$.

Observe

$$p(x) = q(x)^2 \Leftrightarrow p(x) = v(x)^{\mathrm{T}} \left(c(q) c(q)^{\mathrm{T}} \right) v(x)$$

Notice: $c(q)c(q)^{\mathrm{T}} \succeq 0$.

In general, p is SOS if and only if there exists $M \succeq 0$ such that

$$p(x) = v(x)^{\mathrm{T}} M v(x)$$

SDP/SOS/Positive polynomials

Suppose we are interested in

$$p^* = \min p(x)$$

Observe that

$$p^* = \max\{\lambda : p(x) - \lambda \geq 0 \ \ \forall x \in \mathbb{R}^n\}$$

Shor's method: approximate p^* with

$$p_{SOS} := \max\{\lambda : p(x) - \lambda \text{ is SOS}\}$$

Latter can be formulated as an SDP:

$$\begin{array}{ll} \max\limits_{\lambda, M} & \lambda \\ \text{s.t.} & p(x) - \lambda = v(x)^{\mathrm{T}} M v(x) \\ & M \succeq 0 \end{array}$$

The cone of copositive matrices

$$\mathcal{C}_n := \{M \in \mathbb{S}^n : x^{\mathrm{T}} M x \geq 0 \text{ for all } x \geq 0\}$$

Copositive matrices appear in:

complementarity problems

moment problems

combinatorial problems

Example: (Motzkin & Strauss, de Klerk & Pasechnik)
Assume $G = (V, E)$ loopless undirected graph and let $A(G)$ be its incidence matrix. Then

$$\alpha(G) = \min\{\lambda : \lambda(I + A(G)) - ee^{\mathrm{T}} \in \mathcal{C}_n\}$$

Murty & Kabadi: Checking membership in $\mathcal{C}_n$ is NP-hard.

Approximating the copositive cone via SOS

Theorem (Pólya 1928)

If $M \in int(\mathcal{C}_n)$ then for $r \in \mathbb{N}$ sufficiently large

$$\left(\sum_{i=1}^{n} x_i\right)^r x^{\mathrm{T}} M x = \sum_{|\beta|=r+2} c_\beta x^\beta$$

where all c_β are positive.

Parrilo 2001: Based on Pólya's Theorem, construct (via SOS) cones $\mathcal{K}_n^r \subseteq \mathcal{C}_n$ with $\mathcal{K}_n^r \uparrow \mathcal{C}_n$.

Theorem (Zuluaga, Vera, P. 2003)

$M \in \mathcal{K}_n^r$ if and only if

$$\left(\sum_{i=1}^{n} x_i\right)^r x^{\mathrm{T}} M x = \sum_{|\beta| \leq r+2} q_\beta(x) x^\beta$$

where each q_β is SOS.

Zuluaga, Vera, P. 2003: General version of the above for approximating

$$\{p : p(x) \geq 0 \text{ for all } x \in \{x : g_i(x) \geq 0\}\}$$

given $g_1, \ldots, g_k$.

Unifies a number of related results by Lasserre, Kojima, Parrilo,...

SDP descriptions of $\mathcal{K}_n^r$

$$M \in \mathcal{K}_n^r \Leftrightarrow \mathcal{L}(\vec{P}) \leq \mathcal{B}(M) \text{ for some } \vec{P} \succeq 0.$$

$r = 0$: $\vec{P} = P \in \mathbb{S}^n$

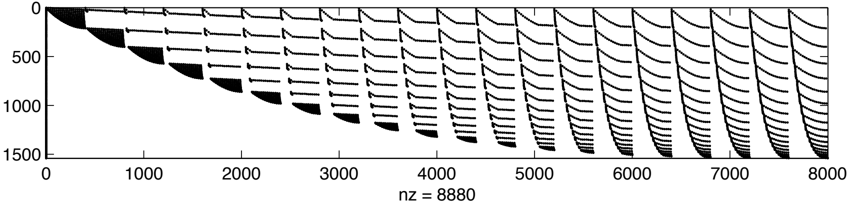
$r = 1$: need n matrices in $\mathbb{S}^n$.

$r = 2$: need $\binom{n}{2}$ matrices in $\mathbb{S}^n$ and one in $\mathbb{S}^{\binom{n}{2}}$.

$$\vdots$$

Structured SDPs

Matrix $\mathcal{L}$ for $r = 1,\ n = 20$



Both $\mathcal{L}$ and $\mathcal{B}$ have very special structure.
Exploit such structure to simplify the SDPs.

Example: Recall

$$\alpha(G) := \min\{\lambda : \lambda(I + A(G)) - ee^{\mathrm{T}} \in \mathcal{C}_n\}$$

De Klerk & Pasechnik: Approximate $\alpha(G)$ by

$$\vartheta^{(r)}(G) := \min\{\lambda : \lambda(I + A(G)) - ee^{\mathrm{T}} \in \mathcal{K}_n^r\}$$

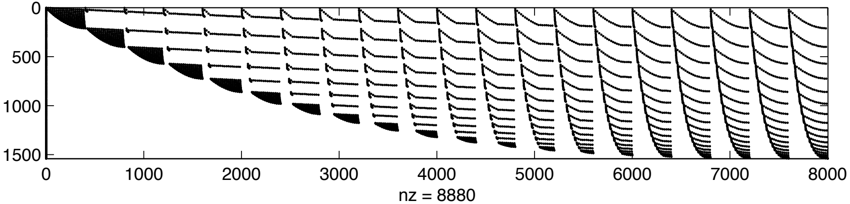
Since $\mathcal{K}_n^r \uparrow \mathcal{C}_n$, we have $\vartheta^{(r)}(G) \downarrow \alpha(G)$.

Structured SDPs: $\vartheta^{(1)}(G)$ for vertex-transitive graphs

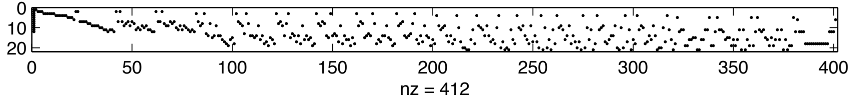
Theorem (Vera & P. 2006)

If G is vertex-transitive under an abelian automorphism group, then $\vartheta^{(1)}(G)$ can be formulated as a reduced SDP in one matrix variable P (instead of n matrix variables).

E.g., for $n = 20$ can reduce from an SDP with constraint matrix



to one with constraint matrix



$\vartheta^{(1)}(G)$ for vertex-transitive G

Want smallest λ such that

$$\sum_{i=1}^{n} x_i x^{\mathrm{T}} \left(\lambda(I + A(G)) - ee^{\mathrm{T}}\right) x = \sum_{i=1}^{n} x_i p_i(x)$$

for $p_i(x) = x^{\mathrm{T}}(P_i + N_i)x, \ P_i \succeq 0, \ N_i \geq 0$

If G vertex transitive can take $p_i(x) = p(\Pi_i x)$.
Π_i: permutation that sends $1 \mapsto i$.
For one single $p(x)$.

Some examples

G	$\alpha(G)$	$\vartheta^{(1)}(G)$	$\vartheta^{(0)}(G)$
C_{2m+1}	m	m	$m + 1/2 - O(1/m)$
C_5^3	10	10.935	11.180
C_7^2	10	10.269	11.006
C_7^3	33	35.341	36.517
C_9^2	17	18.000	19.010
$C_5 \times C_7$	7	7.000	7.4185
$C_5 \times C_7^2$	22	23.853	24.612

Work in progress (Vera & P):

Use these types of reductions to estimate $\alpha(C_7^4), \ldots$

Structured SDPs

SOS approaches often yield SDPs with invariant properties. Get

$$\begin{array}{ll} \min & C \bullet X \\ \text{s.t.} & A_i \bullet X = b_i,\ i = 1, \ldots, m \\ & X \succeq 0. \end{array}$$

Such that for a large multiplicative group G of permutation matrices

$$C = P^{\mathrm{T}} C P,\ A_i = P^{\mathrm{T}} A_i P, \text{ for all } P \in G \text{ and all } i$$

In this case the SDP has a solution in the "centralizer" of G:

$$C(G) := \{X : P^{\mathrm{T}} X P = X\}.$$

Theorem (De Klerk, Pasechnik, Schrijver 2005)

Construct $D_j \in \mathbb{S}^n$, $L_j \in \mathbb{S}^d$, $j = 1, \ldots, d$ *with* $d \leq n$ *such that*

$$X \in C(G),\, X \succeq 0 \Leftrightarrow X = \sum_{j=1}^{d} x_j D_j \text{ and } \sum_{j=1}^{d} x_j L_j \succeq 0$$

Here d: number of G-orbits in $\mathbb{S}^n$.
If G is large then $d << n$.

Limitation:
G has to be large (relative to n).
In particular, it does not apply to $\vartheta^{(1)}(G)$

Structured SDPs

Consider

$$\begin{array}{rl} \min & C \bullet X \\ \text{s.t.} & \mathcal{A}X = b \\ & X \succeq 0. \end{array}$$

Here $\mathcal{A}X := \begin{bmatrix} A_1 \bullet X & \cdots & A_m \bullet X \end{bmatrix}^{\mathrm{T}}$.
Assume $(\mathcal{A}, b, C)$ has the following invariant property:

$$\mathcal{A} = M_g \mathcal{A} N_g^{-1},\ b = M_g b,\ C = N_g^{\mathrm{T}} C$$

for some linear representations $\{M_g : g \in G\}$ and $\{N_g : g \in G\}$ of a finite group G.

Theorem (Vera & P. 2005)

Under the above assumptions can reduce the SDP to a smaller conic program

$$\begin{array}{rl} \min & \langle \bar{c}, x \rangle \\ s.t. & \bar{A}x = \bar{b} \\ & x \in K. \end{array}$$

Reduction is constructive (via numerical linear algebra) and can be incorporated within ipm algorithms.

Example:

$$\begin{bmatrix} a & b & b \\ b & a & b \\ b & b & a \end{bmatrix} \succeq 0 \Leftrightarrow a \geq b \text{ and } a \geq -2b.$$

Can reduce SDP constraint

$$\begin{bmatrix} a & b & b \\ b & a & b \\ b & b & a \end{bmatrix} \succeq 0$$

to an LP constraint

$$\begin{bmatrix} a - b \\ a + 2b \end{bmatrix} \geq 0$$

De Klerk-Pasechnik-Schrijver's approach does not identify this reduction.

Interior-point methods beyond SDP

Heart of ipm: "barrier function".

(LP) Barrier function for $\mathbb{R}^n_{++}$: $-\sum_{i=1}^n \log x_i = -\log x_1 \cdots x_n$

(SDP) Barrier function for $\mathbb{S}^n_{++}$: $-\log\det X$

Consider

$$\begin{array}{rl} \min & \langle c, x \rangle \\ \text{s.t.} & Ax = b \\ & x \in K \end{array}$$

where K is a convex cone.

Can solve via ipm as long as a barrier for $\mathrm{int}(K)$ is available.

Amount of work depends on how complicated the barrier function is.

Symmetric cones

Definition: Assume $K \subseteq \mathbb{R}^n$ is a convex cone.

$$\mathrm{Aut}(K) := \{g \in GL(\mathbb{R}^n) : gK = K\}.$$

Definition: Assume K is a convex cone. K is homogeneous if for all $x, y \in K$ there exists $g \in \mathrm{Aut}(K)$ such that $y = gx$.

Definition: Assume K is a convex cone. K is symmetric if it is homogeneous and $K = K^*$.

Nesterov & Todd 1997: Generalized ipm from SDP to conic programming over symmetric cones. (Currently implemented in `SeDuMi`, `SDPT3`.)

Symmetric cones

Examples of symmetric cones:

(i) Psd $n \times n$ symmetric matrices.

(ii) Lorentz cone: $\{x \in \mathbb{R}^n : x_1 \geq \|(x_2, \dots, x_n)\|\}$.

(iii) Psd $n \times n$ Hermitian matrices with entries in $\mathbb{C}$.

(iv) Psd $n \times n$ Hermitian matrices with entries in $\mathbb{H}$.

(v) Psd 3×3 Hermitian matrices with entries in $\mathbb{O}$.

Theorem

Every symmetric cone is a product of cones of types (i)–(v).

Symmetric cones are slices of $\mathbb{S}^n_+$

Theorem (Hauser & P, 2004)

- *Cones of types (ii) – (v) can be written as* $\{x : Lx \in \mathbb{S}^d_+\}$ *for some explicit L. (Thus have barrier* $x \mapsto -\log\det Lx$*.)*
- *In each case get* $\det(Lx) = p(x)q(x)$ *where factor* $q(x)$ *is "reduntant"*
  Hence barrier $x \mapsto -\log\det Lx$ *can be simplified to* $x \mapsto -\log p(x)$*.*

Example

$$\{x \in \mathbb{R}^n : x_1 \geq \|(x_2, \dots, x_n)\|\} = \left\{ x : \begin{bmatrix} x_1 & x_2 & x_3 & \cdots & x_n \\ x_2 & x_1 & 0 & \cdots & 0 \\ x_3 & 0 & x_1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ x_n & 0 & 0 & \cdots & x_1 \end{bmatrix} \succeq 0 \right\}.$$

$$\det \begin{bmatrix} x_1 & x_2 & x_3 & \cdots & x_n \\ x_2 & x_1 & 0 & \cdots & 0 \\ x_3 & 0 & x_1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ x_n & 0 & 0 & \cdots & x_1 \end{bmatrix} = x_1^{n-1}(x_1^2 - x_2^2 - \cdots - x_n^2)$$

Work in progress...

Hauser, Vera, P: Consider $\{x : Lx \in \mathbb{S}^d_+\}$ for a given L. When can we factor $\det(Lx) = p(x)q(x)$ so that $q(x)$ is redundant?

Special cases:

- Symmetric cones
- Subsets of $\mathbb{S}^d_+$ with a large symmetry group
- More...?

Redundant factor $q(x)$ has to do with some invariant property of the cone $\{x : Lx \in \mathbb{S}^d_+\}$

The polynomial $\det(Lx)$ as well as its factors $p(x), q(x)$ are "hyperbolic" polynomials.

Hyperbolic polynomials

Definition: A polynomial $p(x)$ of degree d is hyperbolic if

- $p(tx) = t^d p(x)$, i.e., p is homogeneous
- there exists e such that for every x the roots of

$$t \mapsto p(x + te)$$

are real.

Examples:

- $x_1 \cdots x_n$, $e = \begin{bmatrix} 1 & \cdots & 1 \end{bmatrix}^{\mathrm{T}}$
- $\det X$, $e = I$,
- $x_1^2 - (x_2^2 + \cdots + x_n^2)$, $e = \begin{bmatrix} 1 & 0 & \cdots & 0 \end{bmatrix}^{\mathrm{T}}$.

SDP/Hyperbolic polynomials

Theorem (Gårding, 1959)

Assume p is hyperbolic. Then each connected component of $\{x : p(x) > 0\}$ is an open convex cone.

Hyperbolicity cone: a component of $\{x : p(x) > 0\}$ for hyperbolic polynomial p.

Theorem (Güler, 1996)

Assume p is hyperbolic. Then $-\log p(x)$ is a barrier function for each component of $\{x : p(x) > 0\}$

LP, SOCP and SDP are special cases of hyperbolicity cones:

> $\mathbb{R}^n_{++}$: component of $\{x : x_1 \cdots x_n > 0\}$ that contains $\begin{bmatrix}1 & \cdots & 1\end{bmatrix}$.
> $\mathbb{S}^n_{++}$: component of $\{X : \det X > 0\}$ that contains I.
> $\mathrm{int}(\mathcal{Q}_n)$: component of $\{x : x_1^2 - (x_2^2 + \cdots + x_n^2) > 0\}$ that contains $\begin{bmatrix}1 & 0 & \cdots & 0\end{bmatrix}^{\mathrm{T}}$.

Lax conjecture, 1958

Original version:

> *A polynomial $p(x, y, z)$ is hyperbolic if and only if there exist $A, B, C \in \mathbb{S}^d$ such that*
>
> $$p(x, y, z) = \det(xA + yB + zC)$$

Theorem (Helton & Vinnikov, 2002)

Lax conjecture is true.

Lewis, Parrilo, Ramana 2003: Observed that Helton & Vinnikov proved Lax conjecture.

General version (still open):

> *Any hyperbolicity cone is a slice of $\mathbb{S}^n_+$.*

Homogeneous cones

Recall: A convex cone K is homogeneous if for all $x, y \in K$ there exists $g \in \text{Aut}(K)$ such that $y = gx$.

Theorem (Güler 1996)

If K homogeneous then K is a hyperbolicity cone.

Theorem (Chua, Faybusovich 2003)

Every homogeneous cone is a slice of $\mathbb{S}^n_+$.

Conjecture (Hauser, P.): Given a hyperbolic polynomial $p(x)$ can construct $q(x)$ and L such that

$$p(x)q(x) = \det Lx$$

for some appropriate L.

Concluding remarks

- SDP approach to problems involving polynomials via SOS. Resulting SDPs are large but highly structured
- Conversely, polynomials suggest extensions of SDP.
- Slices of $\mathbb{S}_+^n$ play a special role: homogeneous & symmetric cones are slices of $\mathbb{S}_+^n$.
- Nice interaction of ideas from optimization/algebra/analysis